(RESEARCH ARTICLE)

Statistical modeling of antifungal activity of substituted benzo-1-thia- and selenium diazoles

Mukhomorov Vladimir K

Università di Napoli Federico II, Via Cintia, I-80126, Napoli, Italy.

Abstract

Statistical methods revealed a close relationship between antifungal activity (suppression of the growth of the mycelium of the fungi of the species *Venturia inaqualis*, *Aspergillus niger* and *Fusarium moniliforme*) and the molecular structure of the substituted benzo-1-thiadiazoles and benzo-1-seleniumdiazoles. The molecular pseudopotential is used as the explanatory variable in the regression equations. It was established the threshold dependence of the antifungal activity on the pseudopotential of the molecule. The explanatory variable exceeding of its threshold value leads to a rapid linear increase in bioactivity for all species of fungi. Antifungal activity of drugs is low to the threshold value of the molecular pseudopotential and does not have a significant relationship with the variability of the molecular structure of the chemical compounds.

1. Introduction

The antifungal activityofbenzo-1-thiadiazoles, which has been detected at various test sites, is well known. [1-3]. However, there is no a unified method for the primary testing of antifungal drugs. [4]. This makes it difficult to identify the quantitative relationship between the chemical structure of compounds and their biological effects. This article analyzes by statistical methods of antifungal activity of the benzo-1-thia- and selenium-diazoles series(suppressing mycelium growth of the fungi of the species *Venturia inaqualis* (Ve), *Aspergillus niger* (As) and *Fusarium moniliforme* (Fu)) [5]. This will allow significant molecular parameters (explanatory variables) of chemical compounds for the purposes of predicting new drugs, within the framework of the heterocycles under study (Fig. 1). In addition, these studies will make some assumptions about the mechanism of biological action of drugs.
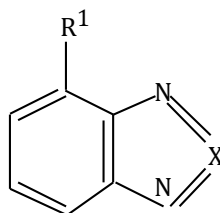

Figure 1 Molecular structure of benzo-1-thia- (X = S) and selenium-diazoles (X = Se).

* Corresponding author
E-mail address: vmukhomorov@mail.ru

The search for the connection between the molecular structure of a chemical compound and its bioactivity will be based on the idea that the objects of study possess some effective electrostatic molecular potential approximated by pseudopotential (1) and (2). Molecular potential can affect the biological system, thereby interfering with the mechanisms regulating the processes of vital activity, that is, determine the biological activity of a chemical compound.

2. Material and methods

The calculation of the real molecular potential is associated with complex quantum chemical calculations, which greatly complicates the construction of a practically convenient mathematical model. At the same time, the pseudopotential method allows one to reliably reproduce many properties of condensed media. Molecular pseudopotential is determined by the sum of the model potentials of the atoms forming the molecule [6].

Here it is proposed to use the average number of electrons on the outer shell of atoms in a molecule for the purpose of identifying the relationship of the biological action of chemical compounds with their molecular structure:

$$Z = \sum_i n_i Z_i / N, \tag{1}$$

here n_i is the number of atoms of the i-th kind; Z_i is the number of electrons in the outer electron shell. Summation is performed over all atoms in the molecule; $\sum_i n_i = N$ is the total number of atoms. Model pseudopotential [6] can be written as follows

$$V(r) = \begin{cases} -Z|e|/r, & r > R_M \\ f(r) + F(r), & r \le R_M, \end{cases} \tag{2}$$

where Z is determined by equation (1); $f(r)$ and $F(r)$ are corrections [6] to the Coulomb potential, which depend on the distance r between the core of the molecule and the electron; there is an electron charge, R_M is the radius of the scattering center. It can be shown that the parameter Z is a common factor for the pseudopotential (2) [7,8].

The method of model molecular pseudopotential assumes that only electrons are taken into account on the outer (valence) shell of the scattering center. It is well known that the chemical properties of molecules are determined by a small group of outer shell electrons. The properties of the remaining electrons have almost no impact on the chemical processes in which the molecule participates. This approximation is sometimes referred to as the "frozen core approach". Valence and core electrons are significantly separated not only by the energy scale, but also spatially. Moreover, changes in one electronic subsystem have little effect on changes in another electronic subsystem. In this approximation, the external electrons do not move in the real Hartree-Fock force field of the molecule, but in a much weaker field of the pseudopotential. The parameter relating to the variation of the potential in molecules is the average number of valence electrons per atom in the molecule. This is certainly an inference that will be used in our research.

In accordance with the pseudopotential model, the average number Z of electrons on the outer electron shells of atoms in a molecule will be used as a molecular trait.

Let us analyze the group of drugs presented in Table 1. This group includes chemical compounds having only one substituent in the R[1] position of the benzene ring.

Table 1 Analysis of group of drugs

N	Substituent R^1	Z	Z_{sub}	Bioactivity [5]			μ, D [1) [10]	MR [9]
				Ve, %	As, %	Fu, %		
Sulfur containing chemical compounds								
1	NO_2	accept.	4.00	5.67	47	21	42	-4.0	7.36
2	SO_2H	accept.	3.88	4.75	38	9	32	-3.8	8.87[2)
3	COOH	accept.	3.75	4.25	18	5	20	-1.0	6.93
4	SH	donor	3.57	3.50	5	5	0	+0.7	9.22
5	OH	donor	3.57	3.50	5	11	7	+1.6	2.85
6	NH_2	donor	3.33	2.75	5	5	7	+1.5	5.42
7	CH_2NH_2	donor	3.24	2.17	12	10	0	-	10.0[2)
8	$CH_2SPO(ONa)_2$	accept.	3.64	3.70	0	0	0	-	20[2)
9	$CH_2NH(CH_2)_2S_2O_3Na$	donor	3.38	3.24	16	6	0	-	40[2)
Selenium containing chemical compounds								
10	NO_2	accept.	4.67	5.67	100	100	65	-4.0	7.36
11	OH	donor	4.29	3.50	23	26	48	+1.6	2.85

[1) Conventionally adopted to direct the dipole moment of the C – R^1 bond: the minus sign is from the substituent to the benzene ring; the plus sign is from the benzene ring to the substituent. [2) The MR index was estimated approximately by an additive scheme.

We first analyze the sample containing only sulfur containing chemical compounds. Now let's check the homogeneity of the samples, for example, for the resultant attribute Fu bioactivity and for the explanatory variable Z:

$N_1 = 9$; $Fu^{av} = 12.00 \pm 5.26\%$, $Fu^{min} = 0$, $Fu^{max} = 42\%$, $S_{Fu} = 15.77$, $\tau_{Fu}^{min} = 0.76 < \tau_{Fu}^{max} = 1.90 < \tau_{0.05}^{cr}(f=9) = 2.35$; $Z^{av} = 3.60 \pm 0.08$, $Z^{min} = 3.24$, $Z^{max} = 4.00$, $S_{Z1} = 0.254$; $\tau_Z^{min} = 1.41 < \tau_Z^{max} = 1.58 < \tau_{0.05}^{cr}(f=9) = 2.35$. (3)

Here S is the standard deviation; Fu^{min} and Fu^{max}, Z^{min} and Z^{max} are the minimum and maximum values of attributes in the samples; Fu^{av} is the mean value of the Fu-activity. Z^{av} is the mean value of the explanatory variable; τ is the Grubbs criterion; f is the number of degrees of freedom. From inequalities (3) it follows that the sets Fu and Z are homogeneous at the 95% confidence level. Using the attribute Z as an explanatory variable, the following rectilinear regression was obtained:

$N_1 = 9$, $Fu(Z) = a_0^{(1)} + a_1^{(1)} \cdot Z$, $R_1 = 0.85 \pm 0.20$, $R_1 > R_{0.05}^{cr}(f=7) = 0.666$ [11], $RMSE_1 = 8.94$, $a_0^{(1)} = -177.5 \pm 44.90$, $a_1^{(1)} = 52.70 \pm 12.46$, $t(a_1^{(1)}) = 4.23 > |t(a_0^{(1)})| = 3.95 > t_{0.05}^{cr}(f=7) = 2.365$, $F = 17.88 > F_{0.05}^{cr}(f_1 = 1; f_2 = 7) = 5.59$.

(4)

From statistics (4) it follows that the regression coefficients are statistically significant at a 95% confidence level. The correlation coefficient and the Fisher criterion cannot be interpreted as a random deviate from zero. Similar statistical relationships can be obtained for the Ve and As bioactivities. Ve and As bioactivities correlate with Fu bioactivity:

$N = 9$, Ve = $b_0^{(1)} + b_1^{(1)} \cdot$Fu, $R = 0.92 \pm 0.15$, $R > R_{0.05}^{cr}(f = 7) = 0.666$, $RMSE = 6.87$, $b_0^{(1)} = 4.96 \pm 2.94$, $b_1^{(1)} = 0.94 \pm 0.15$, $t(b_1)$ = 6.10 > $t_{0.05}^{cr}(f = 7) = 2.365$, $F = 37.19 > F_{0.05}^{cr}(f_1 = 1; f_2 = 7) = 5.59$.

$N = 9$, As = $b_0^{(2)} + b_1^{(2)} \cdot$Fu, $R = 0.75 \pm 0.25$, $R > R_{0.05}^{cr}(f = 7) = 0.666$, $RMSE = 11.38$, $b_0^{(2)} = -0.24 \pm 6.65$, $b_1^{(2)} = 2.06 \pm 0.68$, $t(b_1^{(2)}) = 3.01 > t_{0.05}^{cr}(f = 7) = 2.365$, $F = 9.09 > F_{0.05}^{cr}(f_1 = 1; f_2 = 7) = 5.59$.

(5)

Relationships (4) and (5) suggest that the intensity of the suppression of the fungi mycelium growth *Venturia inaqualis*, *Aspergillus niger* and *Fusarium moniliforme* by chemical compounds depends on the magnitude of the explanatory variable Z and, apparently, is carried out for Fu, Ve and As bioactivities uniformly.

The explanatory variable Z characterizes the molecule as a whole, while the active region of the drug molecule, presumably participating in the interaction with the biosystem, is uncertain. Therefore, it is of interest to perform a regression equation analysis in which the variable Z_{sub} is used as an explanatory variable. The variable Z_{sub} is calculated by the formula (1) for electrons of substituent. For example, the following rectilinear regression was obtained for Fu bioactivity:

$N = 9$,   Fu(Z_{sub}) = $a_0^{(1)} + a_1^{(1)} \cdot Z_{sub}$, $R = 0.88 \pm 0.18$, $R > R_{0.05}^{cr}(f = 7) = 0.666$, $RMSE = 8.07$, $a_0^{(1)} = -37.25 \pm 10.40$, $a_1^{(1)} = 13.22 \pm 2.70$, $t(a_1^{(1)}) = 4.90 > |t(a_0^{(1)})| = 3.59 > t_{0.05}^{cr}(f = 7) = 2.365$, $F = 23.99 > F_{0.05}^{cr}(f_1 = 1; f_2 = 7) = 5.59$.

(6)

The population of elements of the Z_{sub} is homogeneous and normally distributed:

$N = 9$; the value of Wilk-Shapiro statistics is equal to $W_{Zsub} = 0.973 > W_{0.05}^{cr}(f = 9) = 0.829$, $Z_{sub}^{av} = 3.73 \pm 0.35$, $Z_{sub}^{min} = 2.17$, $Z_{sub}^{max} = 5.67$; 95% confidence interval is equal to 2.92 – 4.53, $S_{Zsub} = 1.05$; $\tau_Z^{min} = 1.49 < \tau_Z^{max} = 1.85 < \tau_{0.05}^{cr}(f = 9)$.

(7)

Regression (6) as well as regression (4) are statistically significant. Similarly, we can be obtained the corresponding regressions for the bioactivities of Ve and As. This result indicates that the active region in the molecules is a substituent at the R[1] position of the benzene ring. Apparently, it is the substituent at the R[1] position interacts with the biosystem. Since the value of Z_{sub} characterizes the average number of valence electrons of a substituent, then most likely the area of the biosystem with which the drug interacts has a positive charge.

It is possible in the regression (6) to simultaneously take into account two explanatory variables Z and Z_{sub}. As the analysis showed, this leads to an improvement in the quality of the regression. However, these variables are collinear for the studied series of chemical compounds (Table 1). The correlation coefficient between the variables is equal to $r = 0.98$. It is known [12] that one of the variables should be excluded from the regression equation if the correlation coefficient r is greater than 0.8. One of the variables can be replaced by the difference $Z – Z_{sub}$. Sometimes this transformation allows you to get rid of collinearity. However, collinearity is retained for the chemical compounds of Table 1. For example, the Farrar-Glauber criterion [13] indicates a presence of collinearity between explanatory variables. ($m = 2$):

$$\chi^2 = -(N - 1 - (2m + 5)/6) \cdot \ln(1 - r^2) = 21 > \chi_{0.05}^{2,cr}(f) = 3.84.$$

Here $f = m(m-1)/2$ is the number of degrees of freedom; m is the number of explanatory variables. This criterion can be applied only for a standard normal distribution of residues ($W = 0.902 > W_{0.05}^{cr}(f = 9) = 0.829$). If there is a collinearity between the explanatory variables, the estimates of the regression coefficients can be, firstly, unreliable and, secondly, sensitive to sample data.

For sulfur containing compounds the variation of bioactivities depending on the magnitude of the explanatory variable Z can be approximated by nonlinear regression. Figure 2.A demonstrates the non-linear dependence of the bioactivity Ve on the change in the value of the explanatory variable Z. The nature of the nonlinear change in bioactivity indicates the possible existence of a threshold biological action of drugs. A marked increase in the explanatory variable Z from a value of 3.0 to a value of ∼ 3.65, which can be taken as a threshold value of Z^{thr}, is accompanied, firstly, by a low bioactivity (∼ 10%) of chemical compounds, and, secondly, by a weak variability of bioactivities of Ve, As and Fu. At the same time, if the explanatory variable Z exceeds its threshold value, an intense linear increase in the bioactivity is observed. (Figures 2.B, 2.C and 2.D). A linear regression equation indicates a close relationship between bioactivity and molecular factor Z for the range of $Z \geq Z^{thr}$:

$$N_{Ve} = 4, \, Ve(Z) = a_0^{(Ve)} + a_1^{(Ve)} \cdot Z, \, Z \geq Z^{thr}, \, R = 0.99 \pm 0.11, \, R > R_{0.05}^{cr}(f=2) = 0.95, \, RMSE_{Ve} = 3.92, \, a_0^{(Ve)} = -481.6 \pm 55.32,$$

$$a_1^{(Ve)} = 132.9 \pm 14.48, \, t(a_1^{(Ve)}) = 9.18 > |t(a_0^{(Ve)})| = 8.7 > t_{0.05}^{cr}(f=2) = 4.303, \, F = 84.2 > F_{0.05}^{cr}(f_1=1; f_2=2) = 18.51.$$
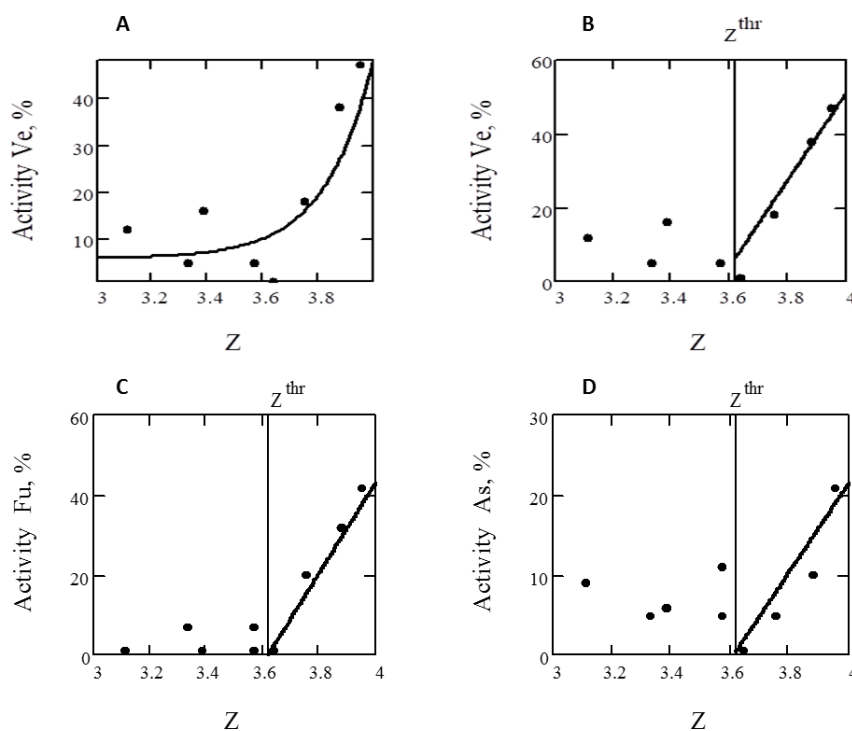(8)

Figure 2 A. Nonlinear relationship of Ve activity with explanatory variable Z: $N = 9$, $Ve(Z) = 6.01 + 3.52 \times 10^{-9} \times \exp(5.81 \cdot Z)$, $RMSE = 8.18$, $R^2 = 0.78$, $F = 31.9 > F_{0.05}^{cr}(f_1=1; f_2=7) = 5.59$. B. The threshold of the Ve activity depending on the magnitude of the explanatory variable Z. C. The threshold of the Fu activity depending on the magnitude of the explanatory variable Z. D. The threshold of the As activity depending on the magnitude of the explanatory variable Z.

Similar threshold dependences take place for bioactivities Fu and As (Fig. 2.C and 2.D):

$$N_{Fu} = 4, \quad Fu(Z) = a_0^{(Fu)} + a_1^{(Fu)} \cdot Z, \quad Z \geq Z^{thr}, \quad R = 0.98 \pm 0.13, \quad R > R_{0.05}^{cr}(f=2) = 0.95, \, RMSE_{Fu} = 4.21, \, a_0^{(Fu)} = -409.7 \pm$$

$$59.40, \, a_1^{(Fu)} = 113.5 \pm 15.55, \, t(a_1^{(Fu)}) = 7.30 > |t(a_0^{(Fe)})| = 6.90 > t_{0.05}^{cr}(f=2) = 4.303, \, F = 53.25 > F_{0.05}^{cr}(f_1=1; f_2=2) = 18.51.$$
(9)

$$N_{As} = 4, \, As(Z) = a_0^{(As)} + a_1^{(As)} \cdot Z, \, Z \geq Z^{thr}, \, R = 0.97 \pm 0.18, \, R > R_{0.05}^{cr}(f=2) = 0.95, \, RMSE_{As} = 2.83, \, a_0^{(As)} = -202.6 \pm$$

$$40.00, \, a_1^{(As)} = 55.37 \pm 10.47, \, t(a_1^{(As)}) = 5.29 > |t(a_0^{(As)})| = 5.07 > t_{0.05}^{cr}(f=2) = 4.303, \, F = 27.96 > F_{0.05}^{cr}(f_1=1; f_2=2) = 18.51.$$
(10)

For small samples ($N < 15$), the best estimate of the correlation coefficient is the ratio [14]:

$$R^* = R \cdot [1 + 0.5 \cdot (1 - R^2)/(N - 3)] . \tag{11}$$

For example, from the relation (11) we obtain the following value $R^* = 0.999$ for the regression (10). All substituents for which $Z \geq Z^{\mathrm{thr}}$ are electron acceptors (sulfur containing compounds).

Using as an explanatory variable Z_{sub} does not change the significance of the detected rectilinear dependencies. (8) – (10). For Fu bioactivity, for example, the following regression equation was obtained (the area $Z \geq Z^{\mathrm{thr}}_{\mathrm{sub}} \equiv Z^{\mathrm{av}}_{\mathrm{sub}}$):

$$N = 4, \ \mathrm{Fu}(Z_{\mathrm{sub}}) = b_0 + b_1 \cdot Z_{\mathrm{sub}}, \ R = 0.96 \pm 0.21, \ R > R^{\mathrm{cr}}_{0.05}(f = 2) = 0.95, \ RMSE = 6.33, \ b_0 = -71.52 \pm 20.3, \ b_1 =$$

$$20.69 \pm 4.37, t(b_1) = 4.74 > t^{\mathrm{cr}}_{0.05}(f = 2) = 4.303, F = 22.46 > F^{\mathrm{cr}}_{0.05}(f_1 = 1; f_2 = 2) = 18.51. \tag{12}$$

This result does not contradict the conclusion (6) that the active center in the molecule is precisely the atom or group of atoms in the position R^1 of the benzene ring. Similar linear regressions can be written for Ve and As activities. It can be noted that for all three bioactivities an intensive linear growth of activity begins after the factor Z reaches the boundary (or threshold) value of $Z = Z^{\mathrm{thr}}_{\mathrm{sub}}$.

Let us check whether the regression dependences Ve(Z) (8) and Fu(Z) (9) differ significantly. Perform a comparison of regressions for these activities. First, we will compare the residual variances:

$$F_{\mathrm{Ve/Fu}} = 4.21^2/3.92^2 = 1.15 < F^{\mathrm{cr}}_{0.05}(f_1 = 2; f_2 = 2) = 19.00. \tag{13}$$

That is, the residual variances do not differ at the 95% confidence level. Further, we compare the regression coefficients that characterize the slope of straight lines:

$$t_{\mathrm{Ve/Fu}} = \frac{a_1^{(\mathrm{Ve})} - a_1^{(\mathrm{Fu})}}{\sqrt{S^2_{\mathrm{Ve/Fu}}\left(\dfrac{1}{(N_{\mathrm{Fu}}-1)S^2_Z} - \dfrac{1}{(N_{\mathrm{Ve}}-1)S^2_Z}\right)}} = 0.91 < t^{\mathrm{cr}}_{0.05}(f = N_{\mathrm{Ve}} + N_{\mathrm{Fu}} - 4) = 2.776. \tag{14}$$

Here we used a summary estimate of residual variances

$$S^2_{\mathrm{Ve/Fu}} = \frac{(N_{\mathrm{Ve}}-2)S^2_{\mathrm{Ve}} + (N_{\mathrm{Fu}}-2)S^2_{\mathrm{Fu}}}{(N_{\mathrm{Ve}} + N_{\mathrm{Fu}} - 4)} = 16.55. \tag{15}$$

The standard deviation $S_Z = 0.156$ was used for the explanatory variable Z. It follows from the inequalities (13) and (14) that the magnitudes of the slopes of the regression lines are statistically indistinguishable. Therefore, the variability of the Ve bioactivity is the same as that of the Fu bioactivity. Similarly, we can quantitatively compare the regressions of Ve(Z) and As(Z), Fu(Z) and As(Z). Apparently, it can be assumed that the mechanism of variability in suppressing of the mycelium growth of the fungi *Venturia inaqualis*, *Aspergillus niger* and *Fusarium moniliforme* with benzothiadiazoles is identical. The differences in the parameters of the regression equations can be attributed to the random fluctuations of the sample data. For forecasting purposes, the linear function is more preferable than a nonlinear dependence in the region of variation of the attribute $Z \geq Z^{\mathrm{thr}}$.

Thus, from the inequalities (13) and (15), it follows that the direct regression lines for the Fu and Ve bioactivities are parallel. Therefore, per unit change in the explanatory variable, the change in the Ve and Fu bioactivities, at a 95% confidence level, is statistically identical.

Now let's check how the regression (4) will change, if we add two selenium-containing compounds to the sample (Table 1). Since statistical comparison is possible only for homogeneous samples, it is first should be checked the uniformity of the set of signs of Z and the set of activity elements, for example, Fu bioactivity:

$N_2 = 11$; $Fu^{av} = 20.09 \pm 6.99$ %, $Fu^{min} = 0$, $Fu^{max} = 65\%$, 95% confidence interval is equal to 4.52 – 35.67 %, $S_{Fu} = 23.18$, $\tau_{Fu}^{min} = 0.87 < \tau_{Fu}^{max} = 1.94 < \tau_{0.05}^{cr}(f = 11) = 2.47$; $Z^{av} = 3.74 \pm 0.14$, $Z^{min} = 3.11$, $Z^{max} = 4.67$; 95% confidence interval is equal to 3.44 – 4.05, $S_{Z2} = 0.449$; $\tau_Z^{min} = 1.40 < \tau_Z^{max} = 2.07 < \tau_{0.05}^{cr}(f = 11) = 2.47$. 　(16)

$N_2 = 11$, $Fu(Z) = a_0^{(2)} + a_1^{(2)} \cdot Z$, $R_2 = 0.96 \pm 0.10$, $R_2 > R_{0.05}^{cr}(f = 9) = 0.602$, $RMSE_2 = 7.28$, $a_0^{(2)} = -179.5 \pm 20.87$, $a_1^{(2)} = 52.93 \pm 5.50$, $t(a_1^{(2)}) = 9.62 > |t(a_0^{(2)})| = 8.60 > t_{0.05}^{cr}(f = 9) = 2.262$, $F = 92.52 > F_{0.05}^{cr}(f_1 = 1; f_2 = 9) = 5.12$. 　(17)

We will compare the regressions (4) and (17) to find out how much the regression changes as the sample volume expands by adding selenium containing chemical compounds (sample volume $N_2 = 11$) compared to the regression for sulfur containing chemical compounds (sample volume $N_1 = 9$). We first compare the residual variances:

$F = 8.94^2 / 7.28^2 = 1.51 < F_{0.05}^{cr}(f_1 = 7; f_2 = 9) = 3.29$. 　(18)

Since $F < F_{0.05}^{cr}(f_1 = N_1 - 2; f_2 = N_2 - 2)$, the null hypothesis on the equality of residual variances at the 95% confidence level is accepted. Consequently, both regression lines are characterized by the same random variance, that is, a similar scattering pattern option around the lines. At the second stage of the test, it is necessary to compare the slopes of the regressions, which are determined by the regression coefficients ($S_{Z1} = 0.254$ (3) and $S_{Z2} = 0.449$ (16)). For this purpose, we use the relations (14) and (15):

$t = 0.016 < t_{0.05}^{cr}(f = N_1 + N_2 - 4) = 2.12$. 　(19)

Consolidate assessment of residual variances is

$$S^2 = \frac{(N_1 - 2)S_1^2 + (N_2 - 2)S_2^2}{N_1 + N_2 - 4} = 64.78.$$ 　(20)

Here $S_1 = 8.94$ (4), $S_2 = 7.28$ (17), $N_1 = 9$, $N_2 = 11$. Since inequality holds (19), it can be recognized that the difference between regression coefficients $a_1^{(1)}$ and $a_1^{(2)}$ is insignificant at a 95% confidence level. We also compare the correlation coefficients $R_1 = 0.85$ and $R_2 = 0.96$. To do this, we use Fisher normalizing transform $z = 0.5 \cdot \ln((1 + R)/(1 - R))$ and the λ-criterion [12]:

$z_1 = 1.26$, $z_2 = 1.95$, $\lambda = |z_1 - z_2|/(1/(N_1 - 3) + 1/(N_2 - 3))^{0.5} = 1.28 < \lambda_{0.05}^{cr} = 1.96$. 　(21)

The inequality (21) implies the absence of a significant difference between the correlation coefficients. That is, there is no significant improvement in regression after an increase in the sample size. A similar analysis can be performed for Ve and As bioactivities.

For samples containing sulfur containing and selenium containing chemical compounds (homogeneous populations), the following significant linear regressions were obtained:

$N_{Ve} = 6$, $Ve(Z) = a_0^{(Ve)} + a_1^{(Ve)} \cdot Z, Z \geq Z^{thr}$, $R = 0.85 \pm 0.27$, $R > R_{0.05}^{cr}(f = 4) = 0.811$, $R^* = 0.89$, $RMSE_{Ve} = 20.67$, $a_0^{(Ve)} = -271.5 \pm 98.03$, $a_1^{(Ve)} = 76.57 \pm 24.19$, $t(a_1^{(Ve)}) = 3.17 > t_{0.05}^{cr}(f = 4) = 2.766$, $F = 10.02 > F_{0.05}^{cr}(f_1 = 1; f_2 = 4) = 7.71$; $Ve^{av} = 37.67 \pm 14.13$ %, $Ve^{min} = 0$, $Ve^{max} = 100\%$, $S_{Ve} = 34.62$, $\tau_{Ve}^{min} = 1.09 < \tau_{Ve}^{max} = 1.80 < \tau_{0.05}^{cr}(f = 6) = 2.07$; $Z^{av} = 4.04 \pm 0.16$, $Z^{min} = 3.64$, $Z^{max} = 4.67$, $S_Z = 0.382$; $\tau_Z^{min} = 1.04 < \tau_Z^{max} = 1.65 < \tau_{0.05}^{av}(f = 6) = 2.07$. 　(22)

$N_{Fu} = 6$, $Fu(Z) = a_0^{(Fu)} + a_1^{(Fu)} \cdot Z$, $Z \geq Z^{thr}$, $R = 0.94 \pm 0.17$, $R > R_{0.05}^{cr}(f = 4) = 0.811$, $R^* = 0.96$, $RMSE_{Fu} = 8.79$, $a_0^{(Fu)} = -190.4 \pm 41.67$, $a_1^{(Fu)} = 55.69 \pm 10.28$, $t(a_1^{(Fu)}) = 5.42 > |t(a_0^{(Fu)})| = 4.57 > t_{0.05}^{cr}(f = 4) = 2.766$, $F = 29.34 > F_{0.05}^{cr}(f_1 = 1; f_2 = 4) = 7.71$; $Fu^{av} = 34.5 \pm 9.27\%$, $Fu^{min} = 0$, $Fu^{max} = 65\%$, $S_{Fu} = 22.70$, $\tau_{Fu}^{max} = 1.34 < \tau_{Fu}^{min} = 1.52 < \tau_{0.05}^{cr}(f = 6) = 2.07$. (23)

$N_{As} = 6$, $As(Z) = a_0^{(As)} + a_1^{(As)} \cdot Z$, $Z \geq Z^{thr}$, $R = 0.92 \pm 0.18$, $R > R_{0.05}^{cr}(f = 4) = 0.811$, $R^* = 0.94$, $RMSE_{As} = 16.45$, $a_0^{(As)} = -334.5 \pm 78.01$, $a_1^{(As)} = 89.35 \pm 19.25$, $t(a_1^{(As)}) = 4.64 > |t(a_0^{(As)})| = 4.29 > t_{0.05}^{cr}(f = 4) = 2.776$, $F = 21.55 > F_{0.05}^{cr}(f_1 = 1; f_2 = 4) = 7.71$; $As^{av} = 26.33 \pm 15.18\%$, $As^{min} = 0$, $As^{max} = 100\%$, $S_{As} = 37.19$, $\tau_{As}^{min} = 0.71 < \tau_{As}^{max} = 1.98 < \tau_{0.05}^{cr}(f = 6) = 2.07$. (24)

Statistical comparison of regressions, for example, Fu bioactivities (12) and (23), leads to the following inequalities for residual variances:

$$F_{Fu/Fu} = 8.79^2/6.33^2 = 1.93 < F_{0.05}^{cr}(f_1 = 4; f_2 = 2) = 19.25. \qquad (25)$$

In accordance with the definitions (14) and (15) the regression coefficients leads to the following inequality:

$$t = 1.13 < \tau_{0.05}^{cr}(f = 6) = 2.447. \qquad (26)$$

The distinction in the correlation coefficients (12) and (23) is also insignificant:

$$z_1 = 1.95, \quad z_2 = 1.74, \quad \lambda = 0.18 < \lambda_{0.05}^{cr} = 1.96. \qquad (27)$$

It follows from the inequalities (25) and (26) that the regressions (12) and (23) differ insignificantly. However, it should be noted that the share of explained variations, determined by the coefficient of determination R^2 (regressions (22) - (24)), is systematically reduced compared to samples containing only sulfur containing diazoles. Moreover, adding only two selenium containing chemical compounds to the samples leads to a significant difference in residual variances for the activity of Ve (regressions (22) and (8) and for the activity of As (regressions (24) and (10)):

$$F_{As/As} = 16.45^2/2.83^2 = 33.8 > F_{Ve/Ve} = 20.67^2/3.92^2 = 27.8 > F_{0.05}^{cr}(f_1 = 4; f_2 = 2) = 19.25. \qquad (28)$$

Adding selenium containing chemical compounds to the sample significantly increases the residual variance (or *RMSE*). Further comparison of regressions is difficult due to the lack of accurate statistical criteria if the conditions (28) set out. Apparently, the condition $Z > Z^{thr}$ is necessary, but not sufficient for the high bioactivity of the chemical compound. It is also important that the substituent is an electron acceptor.

---

3. Results and discussion

A comparative analysis of the relationship between the activity of sulfur containing chemical compounds and selenium containing chemical compounds with the explanatory variable Z leads to the conclusion that benzothiadiazoles and benzoseleniumdiazoles, apparently, belong to various general populations. The common feature of the analyzed drugs and the activities of Ve, As and Fu is the presence of a significant trend, namely the increase of the explanatory variable Z (or the pseudopotential of the molecule), accompanied by an increase in the bioactivity of the chemical compound.

It is possible to note the important property of substitutes which, apparently, on a qualitative level is associated with the manifestation of drug activity. So the highest activity values are characteristic for substituent (Table 1), which have acceptor properties (NO_2, SO_2H, $COOH$). Chemical compounds with donor substituent (SH, OH, NH_2) have a weak bioactivity. Acceptors can be built according to the magnitude of the force of the acceptor influence in the following

row: NO_2 ($Z = 4.0$; $Z_{sub} = 5.67$; $\mu = -4.0$ D) > SO_2H($Z = 3.88$; $Z_{sub} = 4.75$; $\mu = -3.8$ D) > $COOH$($Z = 3.75$; $Z_{sub} = 4.25$; $\mu = -1.0$ D). This series correlates with the bioactivities of Ve, As, and Fu. As is known, the acceptor properties of substituents are characterized by the position of the lowest unoccupied molecular orbital (MO). The lower the position of the level on the energy scale (relative to a molecule without a substituent), the stronger the acceptor properties of the substituent. The donor properties of the substituents are determined by the position of the energy level of the highest occupied MO. The magnitude of the donor influence of the substituents can be arranged in the following row: NH_2 ($Z = 3.33$; $Z_{sub} = 2.33$; $\mu = +1.5$ D) > OH ($Z = 3.57$; $Z_{sub} = 3.50$; $\mu = +1.6$ D) > SH ($Z = 3.57$; $Z_{sub} = 3.50$; $\mu = +0.7$ D). Apparently, the condition $Z > Z^{thr}$ is necessary, but not sufficient for the high bioactivity of the chemical compounds. For high bioactivity of the drug, it is also important that the substituent is an electron acceptor. This result indicates that the active region in the diazole molecules is the substituent at position R^1 of the benzene ring. It could be accepted that the substituent R^1 is the active center of the molecule. Since the value Z_{sub} characterizes the average number of valence electrons of the substituent, it is most likely that the region of the biosystem has a positive charge. This is also indicated by the direction of the dipole moment of the chemical bond $C–R^1$ (see the remark to Table 1).

Table 1 also lists the additional parameters of the substituent: the dipole moment μ of the $C – R^1$ bond [10] and the magnitude of the molar refraction MR [9] characterizing the volume size of the substituent R^1. As analysis has shown, the use of these molecular parameters does not lead to an improvement in the quality of regression equations. This may be due to an insufficient sample size. However, it can be noted, for example, that drugs have a relatively high activity if the three-dimensional size of the R^1 substituent has a value close to $MR_{opt} \sim 7$-9. This range of MR values can be taken as the optimal three-dimensional size of the substituent for the interaction of the drug molecule with the biophase region. The weakening of the activity of Ve, As and Fu is usually accompanied by either a decrease in the amount of molar refraction of the substituent or a marked excess of the MR_{opt} size (Table 1). Such a change in bioactivity may be due to the complementarity of the substituent R^1 size to the local region of the biosystem.

4. Conclusion

A new effective approach to the analysis of the relationship between the antifungal activity of benzo-1-thia and selenium diazoles with their molecular structure is proposed. The method allows to predict the biological activity of chemical compounds when there is no experimental physicochemical information on the properties of chemical compounds. The proposed method allows researchers to perform a quick express analysis of the potential biological activity of new chemical compounds.

References

[1]   Belen'kaya IA, Dyachina ZS and Shulla TA. (1987). Physiologically active substances. Kiev. Issue 19, 56-60.

[2]   Dyachina ZhS, Belen'kaya IA and Prokhorchuk EA. (1987). Physiologically active substances. Kiev. Issue 18, 79-83.

[3]   Van Daalen JJ, Daams J, Koopmen H and Tempel A. (1967). Rec. Trav. Chim. Pays–Bas., 86(2), 1159-1181.

[4]   Golyshin NM. (1970). Fungicides in agriculture. Moscow. 161-176.

[5]   Belen'kaya IA, Dyachina ZS, Mukhomorov VK and Sirik SA. (1991). Structure – Activity. Relation of electronic parameters of substituted benzo-2,1,3-thia and –seleniumdiazoles with their antifungal activity and toxicity. Chemical-Pharmaceutical Journal, 24(12), 49-53.

[6]   Abarenkov IV, Bratsev VF and Tulub AI. (1989). Elementary of quantum chemistry. High School, Moscow. 272-291.

[7]   Husinaga S, Klobukowski M and Sakai Y. (1984). J. Phys. Chem., 88(21), 4880-4886.

[8]   Mukhomorov VK. (2012). Modeling of chemical compounds bioactivity. Relationships of structure - bioactivity. Lambert Academic Publisher, Saarbrücken, Germany. 165.

[9]   Hansch C and Leo A. (1979). Substituent constant for correlation analysis in chemistry and biology. John & Sons, New York - Chichester, Brisbane, Toronto.

[10]   Minkin VI, Osipov OA and Zhdanov YuA. (1968). Dipole moments in organic chemistry. Chemistry, Moscow. 248.

[11]   Sachs L. (1972). Statistische auswertungsmethoden. Springer-Verlag, Berlin, Heidelbereg, New York, 600.

[12]   Förster E and Rönz B. (1979). Methoden der correlations- und regressionsanalyse. Verlag Die Wirtshaft, Berlin. 290.

[13]   Farrar DE and Glauber RR. (1967). Multicollinearity in regression analysis: The problem revised. In: Review of economics and statistics.

[14]   Kobzar AI. (2006). Applied mathematical statistics. For engineers and scientists. FizMatLit, Moscow. 816.

How to cite this article

Mukhomorov VK. (2019). Statistical modeling of antifungal activity of substituted benzo-1-thia- and selenium diazoles. World Journal of Biology Pharmacy and Health Sciences, 1(1), 08-16.